
Casting the Net

Caplan, Priscilla. "To Hel(sinki) and Back for the Dublin Core."
The Public-Access Computer Systems Review 8, no. 4 (1997): 26-30.

Two and a half years ago I reported in this column on the birth of the Dublin Core (DC), a small set of simple, descriptive data elements intended to aid in the discovery of Internet resources (see "You Call It Corn, We Call It Syntax-Independent Metadata for Document-Like Objects," The Public-Access Computer Systems Review 6, no. 4 (1995): 19-23; <URL:http://info.lib.uh.edu/pr/v6/n4/capl6n4.html>). I've just returned from the fifth Dublin Core Workshop in Helsinki and thought this would be a good time for an update. If you haven't been following the Dublin Core lately, you might be surprised.

The seed of the Core was first planted at a W3C (World Wide Web Consortium) conference in the fall of 1994 when some of the attendees collectively wondered whether anything could be done to improve resource discovery on the Web. The end result was a meeting hosted by NCSA (National Center for Supercomputing Applications) and OCLC in March 1995 at which the original Dublin Core data element set was drafted. It rapidly became an international effort with subsequent DC workshops held in Warwick, England, and Canberra, Australia. (If I make this sound like it just happened, it didn't--it was the result of a tremendous effort by a number of dedicated thinkers and organizers, most notably Stu Weibel at OCLC.) There are now more than 700 subscribers to the DC mailing list and the DC home page (<URL:http://purl.oclc.org/metadata/dublin_core/>) lists more than 30 projects in a dozen countries, involving everything from German mathematics papers to Danish government publications.

+ Page 27 +

DC Lite, An Unqualified Success

The Dublin Core itself is a set of 15 (originally 13) simple data elements, such as "title," "contributor," and "date," which are intended to aid in discovery and identification of resources on the network. There are principles governing their use: all elements are optional and all are repeatable. There are optional qualifiers that can be used with any data element: "lang" for the language of the metadata and "scheme" for the authority or standard used in formulating the content (e.g., "scheme = LCSH"). There is also a list of sub-elements being defined that will allow further refinement of the data elements--for example, sub-elements to distinguish different types of dates. Though the Core is inevitably growing in complexity, the base set of 15 simple elements, rather affectionately known as "DC Lite," is reasonably stable and can be used on its own.

The DC gestalt, however, is more than just the data element definitions. There are "crosswalks," or more-or-less official mappings, to and from more complex metadata element sets like

USMARC (U.S. Machine-Readable Cataloging) and GILS (Government Information Locator System). There are a number of canonical representations in various syntaxes, including the simplest and most commonly used, in HTML (Hypertext Markup Language) 2.0. There is also an architecture, the Warwick Framework, for associating different sets of metadata with the same object and for storing metadata either embedded in the object itself or separately. Excitingly, HTML 4.0 contains new META attributes LANG and SCHEME added deliberately to support the Dublin Core. Even more excitingly, the emerging RDF (Resource Description Framework), a standard under development in the W3C, has been heavily influenced by the Warwick Framework and will ultimately provide a means to support it.

The Core Corps

A few of the implementation projects got a chance to report in Helsinki. One of the most positive developments is the emergence of toolkits supporting several levels of Core functionality. One such toolkit is under development by the Nordic Metadata project, designed to support interlibrary loan and resource sharing throughout Scandinavia (<URL:http://renki.helsinki.fi/meta/>). They have templates for data entry, a Z39.50 search system, and converters that can export data in various formats from MARC to ProCite. The software and documentation will be made available in the public domain.

It is interesting to note how many different ways the Dublin Core is being used. Back in 1995 we focused on providing authors with the ability to supply metadata as they mounted their own publications on the Web. This is happening, but not as much as we expected; most DC metadata is being created by catalogers, or by information professionals we wouldn't quite call catalogers, or by other non-authorial agents. Initially we assumed that library catalogers could take advantage of author-created DC metadata to give them a leg up on more complete cataloging, so the earliest crosswalks mapped from DC to more complex standards like USMARC.

Now it appears that an even more common application of DC is as a "lingua franca," a least common denominator for indexing across heterogeneous databases. Say you have some MARC files and some GILS and some EADs (Encoded Archival Descriptions), the simplest way to index them all with some degree of semantic consistency may be to translate them all to DC. As a result we're seeing a need for crosswalks in the other direction--from the more complex sets into the Dublin Core. A more surprising use was described to me by someone from a Swedish project, who noted that they selected Dublin Core not for its simplicity but for its potential complexity. DC and its RDF representation are capable of expressing far more complex linking and hierarchical relationships than MARC.

Un-Finnished Business

The Helsinki workshop was especially useful in identifying issues still to be resolved. Z39.50 works less than splendidly on Dublin Core and there's probably a need for a DC attribute set to supplement Bib-1. RDF is a work in process. Defining a minimal set of sub-elements will be a difficult compromise between utility and complexity, especially since distinctions necessary to one constituency are often irrelevant to another. Guidelines for representing different versions of an object and for representing relationships between objects are needed (two cans of worms that make us librarians squirm).

+ Page 29 +

The most difficult issues are probably logistical and organizational. We like to say that the authority behind the Dublin Core is that of the "emerging consensus." Current implementers, however, would like qualifiers and sub-elements nailed down, and prospective implementers may be waiting until there actually is a standard more solid than the set of documents posted on the DC home page. Some of these documents will become Internet Drafts but more heavyweight standardization through NISO (National Information Standards Organization) or ISO (International Organization for Standardization) may also be worth thinking about. Similarly, if extensions beyond the core Core need to be registered, there must be some registry and system of registration. Standardization and registration both raise issues of who "owns" the Dublin Core, how it changes, and how fast it changes.

Nonetheless, this is a great bandwagon with plenty of opportunities to jump on. If you want to learn more about the Dublin Core, check out the DC home page and keep an eye on D-Lib Magazine (<URL:http://www.dlib.org/>), which provides good coverage of metadata issues. If you or your organization is putting materials on the network without metadata, consider using the Dublin Core for brief descriptions. If you are drafting a metadata standard for a local project, consider basing it on the DC and adding local extensions where necessary. And if you need to use a complex "domain-specific" metadata element set, think about creating crosswalks to and from the Core. It's the Core-ect thing to do.

About the Author

Priscilla Caplan, Assistant Director for Library Systems, University of Chicago Library, 1100 E. 57th Street Chicago, IL 60637. Internet: p-caplan@uchicago.edu.

+ Page 30 +

About the Journal

The World Wide Web home page for The Public-Access Computer Systems Review provides detailed information about the journal and access to all article files:

     <URL:http://info.lib.uh.edu/pacsrev.html>

Copyright